Journal of Medical Virology 57:126–133 (1999)

Primary Structure of Open Reading Frame 2 and 3 of the Hepatitis E Virus Isolated From Morocco†

Jihong Meng,[1,2] Mian-er Cong,[1] Xing Dai,[1,3] Jacques Pillot,[4] Michael A. Purdy,[1] Howard A. Fields,[1] and Yury E. Khudyakov[1]*

[1]*Hepatitis Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia*
[2]*Department of Microbiology and Immunology, Nanjing Railway Medical College, Nanjing, China*
[3]*Institute of Dermatology, Chinese Academy of Medical Sciences, Nanjing, China*
[4]*Unité d'Immunologie Microbienne, Institut Pasteur, Paris, France*

The nucleotide sequence from position 5,014 to 7,186 of the hepatitis E virus (HEV) genome was determined using a set of 10 polymerase chain reaction (PCR) fragments amplified directly from a pool of fecal specimens obtained from patients with well-documented epidemic HEV infection in Morocco. This sequence contains the 3′-terminal region of open reading frame 1 (ORF1), full length ORF2 and ORF3, and a portion of the 3′-noncoding region. The HEV Morocco nucleotide sequence was compared with the corresponding sequences of 13 HEV strains. A region of ORF2 that overlaps with ORF3 was found to be the most conserved region of ORF2, whereas a protein segment encoded by this region was found to be the most variable. Theoretical RNA secondary structure analysis predicted that this region may be folded into a strong secondary structure that may constrain nucleotide sequence variability. In addition, the nucleotide sequence comparison revealed that the HEV Morocco sequence is most homologous to the sequences of the HEV Asian strains compared with the HEV Mexico, swine, and US strains. Phylogenetic analysis performed on the entire ORF2 and ORF3 sequences and on a small fragment of ORF2 allowed classification of the HEV Morocco strain together with a few other known African strains as a separate subtype within the Asian–African genotype. ***J. Med. Virol. 57:126–133, 1999.*** Published 1999 Wiley-Liss, Inc.

KEY WORDS: hepatitis E; nucleotide sequence; structural protein; phylogenetic analysis; heterogeneity

INTRODUCTION

Hepatitis E virus (HEV) is an enterically transmitted agent that causes epidemic and sporadic cases of hepatitis predominantly in developing countries of Asia, Africa and North America [Fields et al., 1993; Purcell, 1996]. The disease generally affects young adults and has a high mortality rate, up to 20%, in infected pregnant women [Khuroo et al., 1981; Datta et al., 1987]. Success in cloning and sequencing of the HEV genome allowed the elucidation of the HEV genetic organization [Reyes et al., 1990; Tam et al., 1991]. The HEV genome is a positive-sense, single-stranded, polyadenylated RNA of approximately 7.5 kb containing three open reading frames (ORFs). ORF1, located at the 5′-end of the genome, is about 5 kb in length and encodes for a putative nonstructural polyprotein that contains motifs characteristic for methyltransferase, papain-like protease, RNA helicase, and RNA-dependent RNA polymerase domains [Koonin et al., 1992]. ORF2 is about 2 kb in length and encodes for the structural protein(s). The small ORF3 of only 369 nucleotides overlaps ORF1 and ORF2 and encodes for a protein of unknown function [Tam et al., 1991].

The genomes of several HEV strains from Asia and North America (Mexico) have been sequenced in their entirety [Tam et al., 1991; Aye et al., 1992a, 1993; Huang et al., 1992; Tsarev et al., 1992; Bi et al., 1993; Yin et al., 1994; Donati et al., 1997]. Partial sequences are also available for other strains from some of these geographic regions [Aye et al., 1992b; Yin et al., 1993; Huang et al., 1995; Panda et al., 1995; Drabick et al., 1997; Gouvea et al., 1997]. Analysis of available HEV sequences revealed significant divergence between Asian and Mexican strains. The nucleotide sequence variability ranged from 1% to 8% among Asian strains to as much as 25% between Mexican and Asian strains [Krawczynski et al., 1997]. A phylogenetic analysis of

*Correspondence to: Yury E. Khudyakov, Hepatitis Branch, Mail stop A-33, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333. E-mail: yek0@ cdc.gov

Accepted 4 September 1998

full length genomes showed the existence of at least two HEV genotypes distributed geographically as Asian genotype and Mexican genotype [Donati et al., 1997]. Recently, a novel swine HEV strain was shown to be closely related to human HEV [Meng, Purcell, et al., 1997]. Sequence analysis of ORF2 and ORF3 of the swine HEV strain suggested that this virus represents the first member of a third genotype. The second member of this genotype was recently isolated from a patient with acute hepatitis in the United States [Schlauder et al., 1998].

Sequence variation between different strains of HEV may present significant problems related to the diagnostic detection of HEV infections and to the development of HEV vaccines. Therefore, it is important to assess the extent and consequence of sequence diversity of different HEV strains from different parts of the world. However, only a few short or isolated nucleotide sequences from African strains of HEV were available for analysis [Chatterjee et al., 1997; van Cuyck-Gandré et al., 1997]. Although more similar to the Asian strains than to the Mexico strain, HEV sequences from Africa were, nonetheless, distinct from all known Asian strains. Recently, we have shown that using an in vitro neutralization assay [Meng, Dubreuil, et al., 1997], the binding of the HEV Morocco strain to the surface of PLC/PRF/5 cells was not efficiently blocked by antibodies against a recombinant C2 protein [Purdy et al., 1992, 1993] that represents the carboxyl two-thirds of the putative capsid protein of the HEV Burma strain, whereas these antibodies were efficient in preventing the binding of the HEV Burma, Mexico, and Pakistan strains [Meng et al., 1998]. The paucity of more HEV African sequence information, especially from ORF2 of the Morocco strain, has hampered analysis of how sequence heterogeneity may contribute to this observation.

In this paper, the nucleotide sequence of full length ORF2 and ORF3 of the HEV Morocco strain is described and compared with 13 corresponding sequences and to 25 small fragments of an ORF2 sequence from different HEV strains available in Genbank. Phylogenetic analysis of these sequences showed that the HEV Morocco strain, together with a few other known African strains, constitute a separate subtype within the Asian genotype of HEV.

MATERIALS AND METHODS

Sample

Fecal specimens collected from 15 patients with well-documented epidemic HEV infection in the Casablanca area of Morocco in 1994 [Benjelloun et al., 1997] were pooled, from which a 10% (wt/vol) suspension in phosphate-buffered saline was prepared. After centrifugation, the supernatant was filtered through a Sterilization Filter Unit (Nalge Company, Rochester, NY) and titrated on PLC/PRF/5 cell culture according to the procedure described earlier [Meng et al., 1996; Meng, Dubreuil, et al., 1997]. The endpoint infectivity titer of this suspension was determined as 10^7 cell culture infectious doses per gram of stool. Aliquots of virus stock were made and stored at –70°C.

Primers

The original sets of nested and semi-nested polymerase chain reaction (PCR) primers were designed using the published sequence of the Burmese HEV prototype strain [Tam et al., 1991] (GenBank accession no. M73218). Sequences of several primers were modified to improve the efficiency of PCR when the corresponding sequences of the Moroccan HEV strain became available during this study. Set 1 contained "external" primers 5′- (4986) CCC TGG GCT CGT TCA TAA CCT GA and 5′- (5419) GGC CTG GTC ACG CCA AGC GGA GC, and "internal" primers 5′- (4986) CCC TGG GCT CGT TCA TAA CCT GA and 5′- (5386) GGC GGG TTG GCG AAC ACG AGG TC. Set 2 contained primers 5′- (5147) ATG CGC CCT CGG CCT ATT TTG TT and 5′- (5455) GGT AGG TCT ACG ACG TGA GGC AAC. Set 3 contained "external" primers 5′- (5360) GCT GGA CCT CGT GTT CGC CAA C and 5′- (5700) TAC TGG GCA TAA TTG GAA GCC TC, and "internal" primers 5′- (5394) TCG GCT CCG CTT GGC GTG ACC A and 5′- (5669) TTA TAT GAG TAT TGG TGC CGT CC. Set 4 contained "external" primers 5′- (5360) GCT GGA CCT CGT GTT CGC CAA C and 5′- (5941) CCC AGA GGT CTC GAC GGA GCG CC, and "internal" primers 5′- (5483) GTC GCT CCG GCC CAT GAC ACC and 5′- (5854) GAC TAA AAT ACG AAC ATC CGT CGA. Set 5 contained "external" primers 5′- (5603) CTG GTT CTT TAT GCC GCC CCT C and 5′- (6214) AGT AAA ATA GAG GTC CTT CAT AA, and "internal" primers 5′- (5711) CGT GCC ACA ATC CGT TAC CGC C and 5′- (6061) CTC AAG GGC AAA GTC CAA CAG CC. Set 6 contained "external" primers 5′- (5966) GGT CTT GTT ATG CTT TGC ATA C and 5′- (6397) CTC TAC AGA TGT ATA CAA CTT AAC, and "internal" primers 5′- (5966) GGT CTT GTT ATG CTT TGC ATA C and 5′- (6214) AGT AAA ATA GAG GTC CTT CAT AA. Set 7 contained "external" primers 5′- (6071) CGC AAC CTT ACC CCC GGT AAC ACC and 5′- (6571) GAG AGA GAG CCA AAG CAC ATC, and "internal" primers 5′- (6152) GAC GGG ACT GCC GAG CTC ACC AC and 5′- (6478) ATC ATA ATC CTG AAT AAC CAC. Set 8 contained "external" primers 5′- (6323) GGC CAG CTG TTC TAC TCC CGT CC and 5′- (6886) AGC GAC CCG GTG CCC GGC GGC A, and "internal" primers 5′- (6407) CAG GAT AAG GGT ATT GCA ATC C and 5′- (6766) CGG CAG GAC AAA GAA GGT CTT C. Set 9 contained "external" primers 5′- (6639) TGA CCT TGG TCA ATG TTG CGA CC and 5′- (7149) GGG GGG CAC AAG CAA ATA AA, and "internal" primers 5′- (6683) TCG CTT GAC TGG ACT AAG GTC AC and 5′- (7129) CTA CAA CTC CCG AGT TTT ACC. Set 10 contained "external" primers 5′- (6639) TGA CCT TGG TCA ATG TTG CGA CC and 5′- (7214) TTT TTT TTT TTT TTT TTT TTC, and "internal" primers 5′- (6981) ATA CCC TGG ATT ACC CTG CTC and 5′- (7208) TTT TTT TTT TTT TTC AGG GAG C. All primers were synthesized at the Biotechnology Core Facility of the National Center

for Infectious Disease, Centers for Disease Control and Prevention, Atlanta, GA.

RNA Extraction and Reverse Transcription

One hundred microliters of the filtered supernatant of the Moroccan strain was mixed with 400 μl of TRIzol reagent (Gibco BRL, Gaithersburg, MD). The mixture was homogenized and incubated for 5 min at room temperature. One hundred microliters of chloroform was added and the mixture was vigorously shaken for 15 sec and incubated at room temperature for 3 min. After centrifugation at 12,000 × *g* for 15 min at 4°C, the aqueous phase was transferred to a fresh microfuge tube. The RNA from the aqueous phase was precipitated by incubating with an equal volume of isopropyl alcohol and 1 μl of glycogen (20 mg/ml) at room temperature for 10 min, and centrifuging at 12,000 × *g* at 4°C for 10 min. After removing the supernatant, the RNA pellet was washed once with 800 μl of 75% ethanol and centrifuged again, dried, and then dissolved in 20 μl of diethylpyrocarbonate treated water. After denaturation for 5 min at 95°C, reverse transcription was carried out at 42°C for 60 min with a total volume of 20 μl containing 50 mM Tris-HCl, 8 mM $MgCl_2$, 30 mM KCl, 1 mM dithiothreithol, 20 units of AMV reverse transcriptase (Boehringer Mannheim, GmbH, Germany), 1 mM of each deoxynucleotide triphosphate (dNTP, Boehringer Mannheim), 20 units of ribonuclease inhibitor (Boehringer Mannhein), 10 pmol of antisense primer, 0.5 μg of randam primer (Promega, Madison, WI), and 10 μl of the denatured RNA. The mixture was heated for 5 min at 95°C, and immediately chilled in ice.

PCR Amplification

The first round PCR amplification was performed with the Expand High Fidelity PCR System (Boehringer Mannheim) in a total volume of 100 μl containing 1× Expand HF buffer containing 1.5 mM $MgCl_2$, 2.6 units of the enzyme mix, 0.2 mM of each dNTP, 100 pM of each sense and anti-sense primers, 5 μl of dimethyl sulfoxide (DMSO), and 10 μl of the denatured cDNA. The first nine fragments were obtained using primer sets 1–9 and the following PCR conditions: denaturation at 94°C for 45 sec, annealing at 60°C for 20 sec, and extension at 72°C for 60 sec for 30 cycles followed by a single 8-min extension step. For fragment 10, the annealing temperature was changed from 60 to 50°C. The second round amplification was performed using 3 μl of the first PCR product and the same amplification conditions as the first round. All PCR products were analyzed by agarose gel electrophoresis.

DNA Purification and Sequencing

PCR products from fragments 1, 3–5, and 7–10 were directly purified with the QIAquick PCR Purification Kit (QIAGEN Inc., Santa Clarita, CA). Fragments 2 and 6 were excised from low melting agarose gel and then purified with the Wizard PCR Preps DNA Purification System (Promega Corporation, Madison WI). The purified DNA was sequenced by the dideoxynucleotide chain termination method using the Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq DNA Polymerase, FS and the 373 DNA sequencer (ABI, Foster City, CA). In all cases, the "internal" primers were used as sequencing primers. Every PCR fragment was sequenced using both strands.

Computer-Assisted Sequence Analysis

Nucleotide and amino acid sequence comparisons were performed with the MegAlign program from the Lasergene software package (DNASTAR Inc., Madison, WI) and the PILEUP program from the Wisconsin Sequence Package (Genetics Computer Group, Madison, WI). Evolutionary distances between pairs of sequences were determined by the DNADIST program from the PHYLIP package (J. Felsenstein, University of Washington, Seattle). Computed distances were used to construct phylogenetic trees by the program NEIGHBOR. RNA secondary structure was predicted using a modified PCFOLD program (M. Zuker, Washington University, St. Louis, MO) [Freier et al., 1986; Jaeger et al., 1989].

Nucleotide Sequence Accession Numbers

The HEV Morocco nucleotide sequence and deduced amino acid sequence were compared with the following corresponding sequences deposited in GenBank: HEV-Algeria [van Cuyck-Gandré et al., 1997] (GenBank accession no. U40046); HEV-Bangladesh [Drabick et al., 1997] (GenBank accession no. AF047864); HEV-Burma-1 [Tam et al., 1991] (GenBank accession no. M73218); HEV-Burma-2 [Aye et al., 1993] (GenBank accession no. D10330); HEV-Burma-3 [Aye et al., 1992b] (GenBank accession no. D10333); HEV-Chad [van Cuyck-Gandré et al., 1997] (GenBank accession no. U62121); HEV-China-1 [Aye et al., 1992a] (GenBank accession no. D11092); HEV-China-2 [Bi et al., 1993] (GenBank accession no. L08816); HEV-China-3 [Yin et al., 1994] (GenBank accession no. L25547); HEV-China-4 [unpublished data] (GenBank accession no. D11093); HEV-China-5 [Aye et al., 1992b] (GenBank accession no. D10332); HEV-India-1 [Panda et al., 1995] (GenBank accession no. U22532); HEV-India-2 [unpublished data] (GenBank accession no. X99441); HEV-India-3 [Donati et al., 1997] (GenBank accession no. X98292); HEV-Mexico [Huang et al., 1992] (GenBank accession no. M74506); HEV-Nepal-1, 2, 3, 4, 5, 6 [Gouvea et al., 1997] (GenBank accession no. AF020492-AF020497); HEV-Pakistan [Tsarev et al., 1992] (GenBank accession no. M80581); HEV-swine [Meng, Purcell, et al., 1997] (GenBank accession no. AF011921); and HEV-US [Schlauder et al., 1998] (GenBank accesion no. AF035437).

Rapid PCR-Based Seroneutralization Assay in Cell Culture

This protocol has been previously described [Meng, Dubreuil, et al., 1997] and details of the cross-neutralization experiments with antibodies against re-

combinant proteins are also published elsewhere [Meng et al., 1998]. Briefly, ~100 cell culture infectious doses of an HEV inoculum were mixed with the serum sample at an appropriate dilution. After incubation at 37°C for 1 hr, the mixture was inoculated onto a monolayer of PLC/PRF/5 cell line. After adsorption for 2 hr at 37°C, the cells were washed with Hanks' solution and RNA was extracted with TRIzol reagent according to the manufacturer's instructions. Reverse transcription (RT)-nested PCR (RT-PCR) was performed by using a set of HEV-PCR primers. The outer primers were YK-1291 (5′-GTT GTC TCA GCC AAT GGC GAG CC) and YK-1294 (5′-GCC TGC GCG CCG GTC GCA ACA), and the inner primers were YK-1292 (5′-TGG AGA ATG CTC AGC AGG ATA A) and YK-1293 (5′-TAA GTG GAC TGG TCG TAC TCG GC). Both the first round and the second round amplifications were carried out according to the following cycling program: denaturation at 94°C for 45 sec, annealing at 60°C for 20 sec, extension at 72°C for 60 sec, and for 30 cycles. Amplicons were separated by agarose gel electrophoresis with size markers and visualized by ethidium bromide fluorescence. Neutralization was determined by the absence of detectable HEV RNA in the inoculated cell culture. A negative serum control, virus control, and uninoculated cell control were always processed for the detection of HEV RNA at the same time for each assay.

RESULTS AND DISCUSSION

Sequence Variations Within the Structural Region of the HEV Genome

The nucleotide sequence from position 5,014 to 7,186 nucleotides (nt) of the HEV genome was determined using a set of 10 PCR fragments amplified directly from a pool of fecal specimens obtained from HEV infected patients residing in Morocco. The nucleotide positioning is in accordance to the prototype HEV Burma strain [Tam et al., 1991]. This sequence of 2,173 nt (GenBank accession No. AF065061) that spans the extreme 3′-terminal region of ORF1, full length ORF2 and ORF3, and a part of the 3′-noncoding region was compared with the corresponding sequences of 13 HEV strains for which full size sequences of ORF2 and ORF3 were available in GenBank. The HEV Morocco nucleotide sequence contains 42 base substitutions that differ from the corresponding sequences of all other HEV strains. The majority of these substitutions were synonymous, with only two substitutions leading to amino acid changes (Fig. 1E). One substitution was a Ser at position 22 where all other HEV strains contained Pro, Thr, or Ala. The another substitution at position 356 was also a Ser in place of Thr found in all other HEV strains. This last substitution was particularly interesting. Recently, it was shown that antibodies obtained against the C2 protein containing the HEV Burma-1 strain ORF2 sequence from 225 aa to 660 aa could efficiently neutralize the infectivity of the Burma, Pakistan and Mexico strains in an in vitro PCR-based neutralization assay. However, these antibodies failed to block the infectivity of the HEV Morocco strain [Meng et al., 1998]. One possible explanation for this observation is that the sequences responsible for the neutralizing epitope(s) of these two strains are different. In support of this explanation, region 225–660 aa of the ORF2 proteins from the HEV Burma-1 strain and the HEV Morocco strain does differ in two positions, 284 and 356 aa. Except for the HEV Burma-1 and China-2 ORF2 proteins, which contain Leu at position 284, all other strains including the Morocco strain contain Pro at this position, suggesting that this residue is unlikely to be responsible for differences observed in the in vitro neutralization assay. However, the Morocco strain contains a unique Ser residue at position 356 compared with the other strains, except for another African strain isolated in Chad [van Cuyck-Gandré et al., 1997]. Sequence analysis performed on the ORF3-encoded protein failed to identify any Morocco strain-specific amino acid substitutions. These findings suggest that Ser in position 356 aa of the ORF2 protein is specific for two African HEV variants and that the specificity of this residue may influence the function of the HEV neutralizing epitope(s).

The HEV Morocco sequence described in this study was obtained from a pool of fecal specimens collected from 15 HEV-infected individuals from a single outbreak of hepatitis E in Morocco (see Materials and Methods). One may argue that this pool contained several HEV strains with some sequences displaying features different from that described above for the Morocco strain. The absence of direct sequencing information on all of the individual isolates composing this pool does not allow strong contradictions of this suggestion. However, first, all of the specimens in the pool were collected from a single outbreak in one location at one time point. Secondly, a PCR fragment comprising the 3′-terminus of the ORF1 and the 5′-termini of the ORFs 2 and 3 has been previously amplified and sequenced from three isolates, F12, F13, and F23, from this pool [Chatterjee et al., 1997]. Two isolates, F13 and F23, displayed identical sequences. The PCR fragment obtained from the third isolate, F12, demonstrated only one nucleotide difference compared with the other two isolates and therefore all three isolates have essentially the same sequence, suggesting no significant heterogeneity between different isolates in this pool. Finally, it is important to emphasize that the HEV Morocco sequence reported in this study was obtained by direct sequencing of PCR fragments. Thus, all features identified in this sequence are common or, at least, predominant for all 15 HEV isolates circulating in this outbreak.

Nonetheless, the lack of efficient neutralization of the Morocco HEV strain found in this pool might be related to the presence of a minor HEV isolate that may have significant sequence differences from the predominant isolates. To address this problem, we sequenced a PCR fragment comprising the region from position 6,393 to 6,600 nt of ORF2 amplified from the HEV adsorbed onto the surface of PLC/PRF/5 cells af-

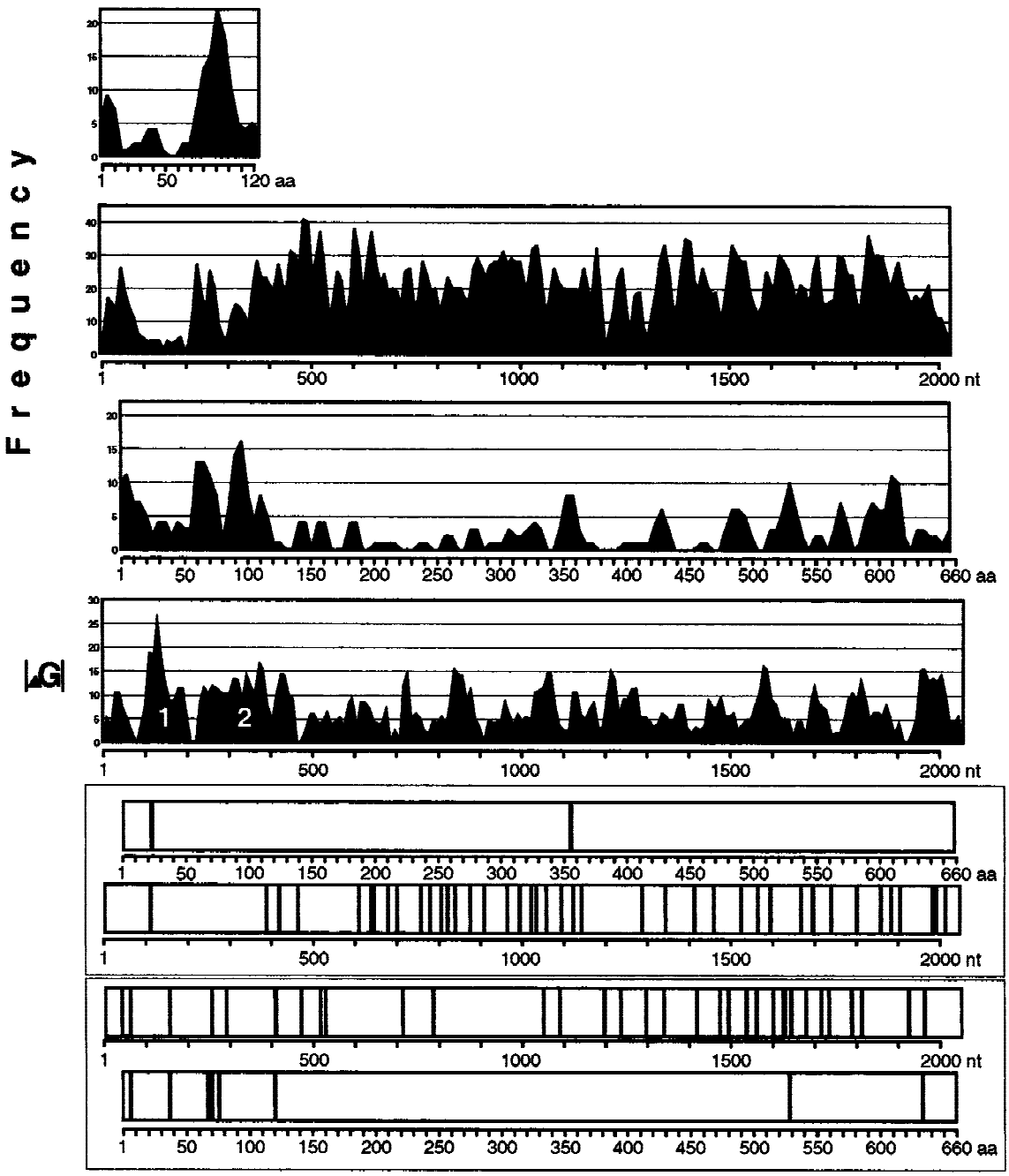

Fig. 1. Sequence variation and RNA secondary structure within open reading frame 2 (ORF2) and ORF3. The distribution of amino acid and nucleotide substitutions was calculated using a sliding window of 10 amino acids or 40 nucleotides, respectively, as the number of changed positions in each of 14 hepatitis E virus (HEV) sequences. **A:** Distribution of amino acid substitutions along the ORF3-encoded protein. **B:** Distribution of the nucleotide substitutions along the RNA strand. **C:** Distribution of the amino acid substitutions along the ORF2-encoded protein. **D:** Distribution of ΔG of RNA secondary structure predicted using a sliding window of 50 nucleotides. Peaks marked 1 and 2 are two structures of strong RNA secondary structure. **E:** Location of amino acid and nucleotide substitutions that are specific for the HEV Morocco strain. **F:** Location of genotype-specific amino acid and nucleotide substitutions.

ter treatment with the neutralizing antibodies against the recombinant C2 protein. Briefly, purified IgG (3.4 mg/ml) obtained by immunization of guinea pigs with the recombinant C2 protein corresponding to a fragment of the HEV Burma ORF2 protein at position of 225–660 aa [Purdy et al., 1992, 1993] were diluted 1:40. Approximately 100 cell culture infectious doses of the HEV Morocco inoculum were incubated with this purified IgG and subsequently inoculated onto a cell monolayer. Similar to our previous observation [Meng et al., 1998], no neutralization activity was detected. The amplicon obtained from inoculated cells was sequenced directly and compared with the corresponding region of the same inoculum without treatment with neutraliz-

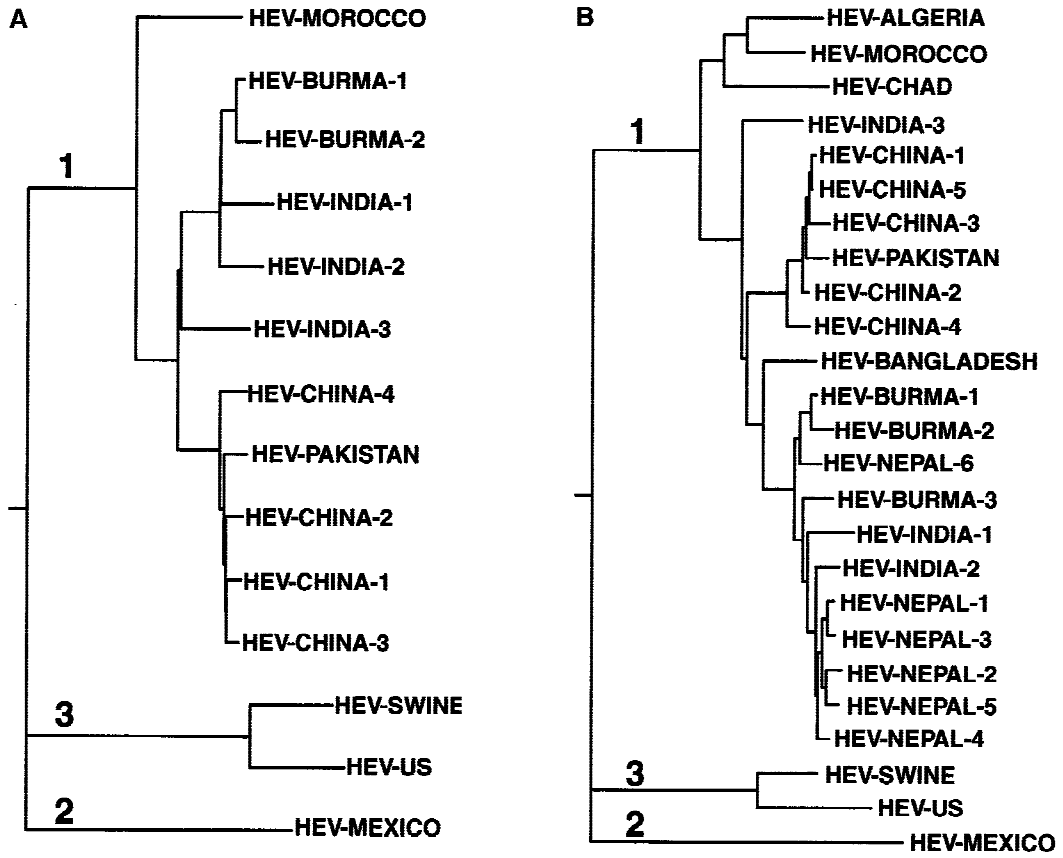

Fig. 2. Proposed phylogenetic trees using (**A**) a 2.0-kb sequences containing the entire open reading frame 2 (ORF2) and ORF3, and (**B**) a short fragment of 424 nucleotides (6,677–7,100 nucleotides) from ORF2.

ing antibodies. No nucleotide differences were found. Therefore, the viruses adsorbed onto the surface of PLC/PRF/5 cells after treatment with the neutralizing antibodies share common sequence features characteristic for the predominant HEV isolates in this pool. Collectively, these data suggest that the HEV population in the Morocco pool of fecal specimens is homogeneous. Thus, the lack of neutralization of the HEV Morocco strain cannot be attributed to significant sequence heterogeneity of isolates found in the pool.

Distribution of Nucleotide and Amino Acid Changes Along the Strand

Additional analysis on the distribution of changes within nucleotide and amino acid sequences of 14 HEV strains was performed. The results are presented in Figure 1A–C. The nucleotide changes were scattered almost uniformly along the entire ORF2, except for a region that overlapped with ORF3 (Fig. 1B). This region may be constrained to nucleotide changes to preserve the functional activity of the ORF2 and ORF3 proteins. However, it is important to note that the most heterogeneous sequence within the ORF2 protein was encoded by the same region (Fig. 1C). In addition, the ORF3-encoded protein was also notably heterogeneous, especially at the C-terminus (Fig. 1A). Thus, despite the lower nucleotide sequence heterogeneity of the 5′-terminal part of the ORF2, the N-terminal region of the ORF2-encoded protein and the C-terminal part of the ORF3-encoded protein were the most variable parts of these proteins.

One possible factor that may constrain the number of nucleotide changes in this region is the secondary structure of the HEV RNA. To examine this possibility, an analysis of RNA secondary structure was performed. The stability of RNA secondary structure, which was generated by using a sliding window of 50 nt, was studied along the entire ORF2 (Fig. 1D). This analysis revealed that the ORF2 and ORF3 overlapping region may be folded into a strong secondary structure, with structure 1 containing the most stable secondary structure and structure 2 containing the most extended strong secondary structure across the entire ORF2 (Fig. 1D). Whereas structure 1 is located within the most conserved region, the region where structure 2 is located is variable, resulting in amino acid changes of the ORF2 and ORF3 encoded-proteins (Fig. 1A–C). These data suggest that mutations within structure 2 occur to retain some elements of the RNA secondary structure, rather than to prevent changes in the protein primary structure. Thus, the RNA structure may serve as the most important constraining fac-

TABLE I. Genotype-Specific Amino Acids Within the HEV ORF2 Protein

Origin of HEV strains	GenBank accession no.	Genotype	Genotype-specific amino acid at the position							
			6	37	67	70	76	120	527	632
Burma-1	M73218	1	I	S	T	A	V	P	S	P
Burma-2	D10330	1	I	S	T	A	V	P	S	P
India-1	U22532	1	I	S	T	A	V	P	S	P
India-2	X99441	1	I	S	T	A	V	P	S	P
India-3	X98292	1	I	S	T	A	V	P	S	P
China-1	D11092	1	I	S	T	A	V	P	S	P
China-2	L08816	1	I	S	T	A	V	P	S	P
China-3	L25547	1	I	S	T	A	V	P	S	P
China-4	D11093	1	I	S	T	A	V	P	S	P
Pakistan	M80581	1	I	S	T	A	V	P	S	P
Morocco	AF065061	1	I	S	T	A	V	P	S	P
Mexico	M74506	2	L	T	A	S	L	S	P	A
Swine	AF011921	3	V	A	V	P	P	A	T	T
US	AF035437	3	V	A	V	P	P	A	T	T

HEV, hepatitis E virus; ORF 2, open reading frame 2.

tor that affects variability of the nucleotide sequence within this region of the HEV genome. This suggestion implies that this RNA secondary structure plays an important role in the HEV replication cycle. Furthermore, it is conceivable that this RNA secondary structure may be involved in the regulation of ORF2 and/or ORF3 protein synthesis by affecting the synthesis of mRNA or the initiation of translation of these proteins.

HEV Genetic Heterogeneity

ORF2 and ORF3 nucleotide sequences of the HEV Morocco strain obtained in this study were compared with the corresponding regions of 13 HEV strains from different parts of the world. The comparison revealed that the HEV Morocco strain sequence is most similar to sequences of the HEV Asian strains (89.3–91.1% similar). The Morocco strain, however, differs from the Asian strains more than the Asian strains differ from each other (92.2–98.8% similar). The nucleotide sequences of the HEV Mexico, swine, and US strains share a lower similarity with the HEV Morocco strain sequence: 80.1%, 78.5%, and 77.9%, respectively. Thus, the data obtained in this study confirm previous observations that African strains of HEV are distinct from other geographic HEV strains, although more similar to Asian strains than to the Mexico strain [Chatterjee et al., 1997; van Cuyck-Gandré et al., 1997]. This conclusion is supported by a phylogenetic analysis performed on the entire ORF2 and ORF3 sequences (Fig. 2A) and on a small fragment of 424 nucleotides (6,677–7,100 nt) of the ORF2 region (Fig. 2B). Both phylogenetic trees identify three HEV genotypes. When the entire ORF2 and ORF3 sequences were analyzed (Fig. 2A), the HEV Morocco strain was grouped together with the Asian strains as a separate branch. When a small fragment of ORF2 was analyzed (Fig. 2B), the Morocco strain clustered together with the other known African strains. This observation suggests that the African strains of HEV represent a separate subtype within genotype 1. The trees in Figure 2 show the existence of two other subtypes: one consisting of HEV strains from China and Pakistan, and the other from Burma, India, and Nepal. The only discrepant findings between the two trees is the positioning of the HEV-India-3 strain, making its phylogenetic relationship within the Asian group of HEV strains uncertain.

Using the results of the phylogenetic analysis and the sequence alignment of all 14 HEV sequences, we have identified 34 nucleotide positions and 8 amino acid positions within ORF2 that are specifically associated with genotypes (Table I). All genotype specific amino acid positions are located within an approximate 150 aa region at the N- and C-termini of the ORF2-encoded protein (Fig. 1F). The genotype-specific nucleotide positions are distributed more uniformly along the sequence (Fig. 1F), although the 3′-terminal half contains the majority of all these genotype-specific positions. Because of the strong RNA secondary structure at the 5′-end of ORF2 (Fig. 1D), this region cannot be amplified efficiently by PCR, rendering the 3′-terminal half of ORF2 the most suitable region for phylogenetic analysis of the HEV genome.

REFERENCES

Aye TT, Uchida T, Ma X, Iida F, Shikata T, Ichikawa M, Rikihisa T, Win KM. 1993. Sequence and gene structure of the hepatitis E virus isolated from Myanmar. Virus Genes 7:95–110.

Aye TT, Uchida T, Ma X-Z, Iida F, Shikata T, Zhuang H, Win KM. 1992a. Complete nucleotide sequence of a hepatitis E virus isolated from the Xinjiang epidemic (1986–1988) of China. Nucl Acids Res 20:3512–3512.

Aye TT, Uchida T, Ma X-Z, Iida F, Shikata T, Zhuang H, Win KM. 1992b. Sequence comparison of the capsid region of hepatitis E virus isolated from Myanmar and China. Microbiol Immunol 36: 615–621.

Benjelloun S, Bahbouhi B, Bouchrit N, Cherkaoui L, Hda N, Mahjour J, Benslimane A. 1997. Seroepidemiological study of an acute hepatitis E outbreak in Morocco. Res Virol 148:279–287.

Bi S-L, Purdy MA, McCaustland KA, Margolis HS, Bradley DW. 1993. The sequence of hepatitis E virus isolated directly from a single source during an outbreak in China. Virus Res 28:233–247.

Chatterjee R, Tsarev S, Pillot J, Coursaget P, Emerson SU, Purcell RH. 1997. African strains of hepatitis E virus that are distinct from Asian strains. J Med Virol 53:139–144.

Datta R, Panda SK, Tandon BN, Madnagopalan N, Bose SL, Acharya SK, Nayak NC. 1987. Acute sporadic non-A, non-B viral hepatitis

of adults in India: epidemiological and immunological studies. J Gastroenterol Hepatol 2:333–345.

Donati MC, Fagan EA, Harrison TJ. 1997. Sequence analysis of full length HEV clones derived directly from human liver in fulminant hepatitis E. In: Rizzetto M, Purcell RH, Gerin JL, Verme G, editors. Viral hepatitis and liver disease. Turin, Italy: Edizioni Minerva Medica. p 313–316.

Drabick JJ, Gambel JM, Gouvea VS, Caudill JD, Sun W, Hoke Jr. CH, Innis BL. 1997. A cluster of acute hepatitis E infection in united nations Bangladeshi peacekeepers in Haiti. Am J Trop Med Hygiene 57:449–454.

Fields HA, Favorov MO, Margolis HS. 1993. The hepatitis E virus: a review. J Clin Immunoassay 16:215–223.

Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci USA 83:9373–9377.

Gouvea V, Snellings N, Cohen SJ, Warren RL, Myint KSA, Shrestha MP, Vaughn DW, Hoke Jr. CH, Innis BL. 1997. Hepatitis E virus in Nepal: similarities with the Burmese and Indian variants. Virus Res 52:87–96.

Huang C-C, Nguyen D, Fernandez J, Yun KY, Fry KE, Bradley DW, Tam AW, Reyes GR. 1992. Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus (HEV). Virology 191:550–558.

Huang R, Nakazono N, Ishii K, Kawamata O, Kawaguchi R, Tsukada Y. 1995. Existing variations on the gene structure of hepatitis E virus strains from some regions of China. J Med Virol 47:303–308.

Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 86:7706–7710.

Khuroo MS, Teli MR, Skidmore S, Sofi MA, Khuroo MI. 1981. Incidence and severity of viral hepatitis in pregnancy. Am J Med 70: 252–255.

Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. Proc Natl Acad Sci USA 89:8259–8263.

Krawczynski K, Mast EE, Purdy MA. 1997. Hepatitis E: an overview. In: Rizzetto M, Purcell RH, Gerin JL, Verme G, editors. Viral hepatitis and liver disease. Turin: Edizioni Minerva Medica. p 305–312.

Meng J, Dubreuil P, Pillot J. 1997. A new PCR-based seroneutralization assay in cell culture for diagnosis of hepatitis E. J Clin Microbiol 35:1373–1377.

Meng J, Guinet R, Pillot J. 1996. Infection of PLC/PRF/5 cells with the hepatitis E virus. In: Buisson Y, Coursaget P, Kane M, editors. Enterically-transmitted hepatitis viruses. Joué-lès-Tours, France: La Simarre. p 336–345.

Meng J, Pillot J, Dai X, Fields HA, Khudyakov Y. 1998. Neutralization of different geographic strains of the hepatitis E virus with anti-hepatitis E virus-positive serum samples obtained from different sources. Virology 249:316–324.

Meng X-J, Purcell RH, Halbur PG, Lehman JR, Webb DM, Tsareva TS, Haynes JS, Thacker BJ, Emerson SU. 1997. A novel virus in swine is closely related to the human hepatitis E virus. Proc Natl Acad Sci USA 94:9860–9865.

Panda SK, Nanda SK, Zafrullah M, Ansari IH, Ozdener MH, Jameel S. 1995. An Indian strain of hepatitis E virus (HEV): cloning, sequence, and expression of structural region and antibody responses in sera from individuals from an area of high-level HEV endemicity. J Clin Microbiol 33:2653–2659.

Purcell RH. 1996. Hepatitis E virus. In: Fields BN, Knipe DM, Howley PM, Chanock RM, Melnick JL, Monath TP, Roizman B, Straus SE, editors. Fields virology, 3rd ed. Philadelphia: Lippincott–Raven. p 2831–2843.

Purdy MA, McCaustland KA, Krawczynski K, Spelbring J, Reyes GR, Bradley DW. 1993. Preliminary evidence that a trpE-HEV fusion protein protects cynomolgus macaques against challenge with wild-type hepatitis E virus (HEV). J Med Virol 41: 90–94.

Purdy MA, McCaustland KA, Krawczynski K, Tam A, Beach MJ, Tassopoulos NC, Reyes GR, Bradley DW. 1992. Expression of a hepatitis E virus (HEV)-trpE fusion protein containing epitopes recognized by antibodies in sera from human cases and experimentally infected primates. Arch Virol 123:335–349.

Reyes GR, Purdy MA, Kim JP, Luk KC, Young LM, Fry KE, Bradley DW. 1990. Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. Science 247:1335–1339.

Schlauder GG, Dawson GJ, Erker JC, Kwo PY, Knigge MF, Smalley DL, Rosenblatt JE, Desai SM, Mushahwar IK. 1998. The sequence and phylogenetic analysis of a novel hepatitis E virus isolated from a patient with acute hepatitis reported in the United States. J Gen Virol 79:447–456.

Tam AW, Smith MM, Guerra ME, Huang C-C, Bradley DW, Fry KE, Reyes GR. 1991. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. Virology 185:120–131.

Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell RH. 1992. Characterization of a prototype strain of hepatitis E virus. Proc Natl Acad Sci USA 89:559–563.

van Cuyck-Gandré H, Zhang HY, Tsarev SA, Clements NJ, Cohen SJ, Caudill JD, Buisson Y, Coursaget P, Warren RL, Longer CF. 1997. Characterization of Hepatitis E virus (HEV) from Algeria and Chad by partial genome sequence. J Med Virol 53:340–347.

Yin S, Purcell RH, Emerson SU. 1994. A new Chinese isolate of hepatitis E virus: comparison with strains recovered from different geographical regions. Virus Genes 9:23–32.

Yin S, Tsarev SA, Purcell RH, Emerson SU. 1993. Partial Sequence comparison of eight new Chinese strains of hepatitis E virus suggests the genome sequence is relatively stable. J Med Virol 41: 230–241.